جامعة ستانفورد: تطبيق "شات جي بي تي" يزداد غباء مع مرور الوقت

توصلت دراسة جديدة أجراها باحثون من جامعتي ستانفورد وبيركلي الأمريكيتين إلى أن نموذج اللغة الكبير (جي بي تي) GPT، الذي يشكل أساس روبوت الدردشة (شات جي بي تي) ChatGPT من شركة (أوبن أي آي) OpenAI أصبح أغبى، وذلك بخلاف ما يدعيه مسؤولو الشركة.

و وجد الباحثون في الدراسة،أنه على مدار بضعة أشهر، غيّر كل من نموذجي (جي بي تي-3.5) و(جي بي تي-4) "سلوكهما" على نحو كبير، إذ أخذت دقة ردودهما تنخفض.
وهذا يؤكد صحة تندر المستخدمين بشأن التدهور الواضح في أداء أحدث إصدارات النموذج في الأشهر التي تلت إصداره.

و كتب الباحثون في ملخص ورقتهم البحثية أن نموذج (جي بي تي-4) في شهر آذار/مارس 2023 كان جيدا جدا في تحديد الأعداد الأولية، وذلك بدقة بلغت نسبتها 97.6%.

و لكن (جي بي تي-4) في شهرحزيران/ يونيو 2023 كان ضعيفا جدا في الأسئلة نفسها مع دقة لم تتعد

2.4%.

معاناة من أخطاء التنسيق

وأضاف الباحثون: "عانى كل من (جي بي تي-4) و(جي بي تي-3.5) المزيد من أخطاء التنسيق في إنشاء التعليمات البرمجية في شهرحزيران/ يونيو مقارنة بشهر آذار/مارس.

أذكى من الإصدار السابق

وتؤكد هذه الدراسة ما قاله المستخدمون منذ أكثر من شهر، وذلك أنهم استخدموا روبوت الدردشة (شات جي بي تي) ChatGPT الذي يعتمد على نموذجي (جي بي تي-3.5) و(جي بي تي-4) فلاحظوا مع مرور الوقت أنه أصبح أغبى.

وأصبح التدهور الواضح في دقته مزعجًا للغاية لدرجة أن نائب رئيس المنتجات لدى (أوبن أي آي)، بيتر ويلندر حاول تبديد الشائعات القائلة بأن التغيير كان مقصودًا.

ونشر ويلندر الأسبوع الماضي تغريدة في موقع تويتر قال فيها: "لا، لم نجعل (جي بي تي-4) أغبى".

وأضاف: "على العكس تمامًا، فنحن نجعل كل إصدار جديد أذكى من الإصدار السابق".

ولفت إلى أن التغييرات في تجربة المستخدم قد تكون ناتجة عن الاستخدام المستمر، قائلًا إنه قد يكون "عند استخدامك (شات جي بي تي) بكثرة، فقد تجد أنك تلاحظ المشكلات التي لم ترها من قبل".

ومع ذلك، فإن أبحاث جامعتي ستانفورد وبيركلي قد تكون أكثر إقناعًا في تفنيد هذه الفرضية.

و في حين أن الباحثين لا يطرحون أسبابًا لحدوث هذا التدهور في الدقة والقدرة، فقد أشاروا إلى أن هذا التدهور الواضح بمرور الوقت يتحدى إصرار شركة (أوبن أي آي) على أن نماذجها تتحسن.

وجاء في الورقة البحثية "وجدنا أن أداء كل من (جي بي تي-3.5) و(جي بي تي-4) وسلوكهما يختلفان اختلافًا كبيرًا فيما بينهما، وأن أداءهما في بعض المهام قد ازداد سوءًا بمرور الوقت".

مضيفة أنه من "المثير للاهتمام"التساؤل عن كون (جي بي تي-4) أصبح أقوى حقًا.

وكتب الباحثون: "من المهم معرفة حقيقة أن التحديثات على النموذج التي تهدف إلى تحسين بعض الجوانب تضر بالفعل بقدرته في أبعاد أخرى".